

What is claimed is:

1 1. A method of displaying text information corresponding to a speech
2 portion of audio signals of a television program to as a closed caption on an video display
3 device, the method comprising the steps of:

4 decoding the audio signals of the television program;

5 filtering the audio signals to extract the speech portion;

6 parsing the speech portion into discrete speech components in accordance with
7 a speech model and grouping the parsed speech components;

8 identifying words in a database corresponding to the grouped speech
9 components; and

10 converting the identified words into text data for display on the display device
11 as the closed caption.

1 2. A method according to claim 1, wherein the step of filtering the audio
2 signals is performed concurrently with the step of decoding of later-occurring audio signals
3 of the television program and step of parsing of earlier occurring speech signals of the
4 television program.

1 3. A method according to claim 1, wherein the step of parsing the speech
2 portion into discrete speech components includes the step of employing a speaker
3 independent model to provide individual words as the parsed speech components.

1 4. A method according to claim 1 further including the step of formatting
2 the text data into lines of text data for display in a closed caption area of the display device.

1 5. A method according to claim 1, wherein the step of parsing the speech
2 portion into discrete speech components includes the step of employing a speaker dependent
3 model to provide phonemes as the parsed speech components.

TOP SECRET

1 6. A method according to claim 5, wherein the speaker dependent model
2 employs a hidden Markov model and the method further comprises the steps of:

3 receiving a training text as a part of the television signal, the training text
4 corresponding to a part of the speech portion of the audio signals;

5 updating the hidden Markov model based on the training text and the part of
6 the speech portion of the audio signals corresponding to the training text; and

7 applying the updated hidden Markov model to parse the speech portion of the
8 audio signals to provide the phonemes.

1 7. A method of displaying text information corresponding to a speech portion
2 of audio signals of a television program to as a closed caption on an video display device, the
3 method comprising the steps of:

4 decoding the audio signals of the television program;

5 filtering the audio signals to extract the speech portion;

6 receiving a training text as a part of the television signal, the training text
7 corresponding to a part of the speech portion of the audio signals;

8 generating a hidden Markov model from the training text and the part of the
9 speech portion of the audio signals;

10 parsing the audio speech signals into phonemes based on the generated Hidden
11 Markov model;

12 identifying words in a database corresponding to grouped phonemes; and

13 converting the identified words into text data for presentation on the display of
14 the audio-visual device as closed captioned textual data.

TOP SECRET

1 8. A method according to claim 7, wherein the step of filtering the audio
2 signals is performed concurrently with the step of decoding of later-occurring audio signals
3 of the television program and step of parsing of earlier occurring speech signals of the
4 television program.

1 9. A method according to claim 7 further including the step of formatting
2 the text data into lines of text data for display in a closed caption area of the display device.

1 10. A method according to claim 7, further comprising the step of
2 providing respective audio speech signals and training texts for each speaker of a plurality of
3 speakers on the television program.

1 11. Apparatus for displaying text information corresponding to a speech
2 portion of audio signals of a television program to as a closed caption on an video display
3 device, the method comprising:

4 a decoder which separates the audio signals from the television program
5 signals;

6 a speech filter which identifies portions of the audio signals that include speech
7 components and separates the identified speech component signals from the audio signals;

8 a phoneme generator which parses the speech portion into phonemes in
9 accordance with a speech model;

10 a database of words, each word being identified as corresponding to a discrete
11 set of phonemes;

12 a word matcher which groups the phonemes provided by the phoneme
13 generator and identifies words in the database corresponding to the grouped phonemes; and

14 a formatting processor that converts the identified words into text data for
15 display on the display device as the closed caption.

FOI 2004-032860

1
2
1
2
1
2
3
4
5
6
7
8
1
2
3
4
5
6
7
8

1
2
1
2
3
4
5
6
7
8
1
2
3
4
5
6
7
8

1
2
3
4
5
6
7
8
1
2
3
4
5
6
7
8

3
4
5
6
7
8
1
2
3
4
5
6
7
8

5
6
7
8
1
2
3
4
5
6
7
8

7
8
1
2
3
4
5
6
7
8

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

5
6
7
8

6
7
8

7
8

